Title
Force for ancient and recent life: viral and stem-loop RNA consortia promote life.

Journal
Annals of the New York Academy of Sciences, 1341(1)

ISSN
0077-8923

Author
Villarreal, Luis P

Publication Date
2015-04-01

DOI
10.1111/nyas.12565

Peer reviewed

Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Issue: *DNA Habitats and Their RNA Inhabitants*

Force for ancient and recent life: viral and stem-loop RNA consortia promote life

Luis P. Villarreal

Center for Virus Research, University of California, Irvine, California

Address for correspondence: Luis P. Villarreal, Department of Molecular Biology and Biochemistry, University of California, 2208 Bio. Sci III, Irvine, CA 92697. lpvillar@uci.edu

Lytic viruses were thought to kill the most numerous host (i.e., kill the winner). But persisting viruses/defectives can also protect against viruses, especially in a ubiquitous virosphere. In 1991, Yarmolinsky *et al.* discovered the addiction modules of P1 phage, in which opposing toxic and protective functions stabilize persistence. Subsequently, I proposed that lytic and persisting cryptic virus also provide addiction modules that promote group identity. In eukaryotes (and the RNA world), a distinct RNA virus–host relationship exists. Retrovirurses/retroposons are major contributors to eukaryotic genomes. Eukaryotic complexity appears to be mostly mediated by regulatory complexity involving noncoding retroposon-derived RNA. RNA viruses evolve via quasispecies, which contain cooperating, minority, and even opposing RNA types. Quasispecies can also demonstrate group preclusion (e.g., hepatitis C). Stem-loop RNA domains are found in long terminal repeats (and viral RNA) and mediate viral regulation/identity. Thus, stem-loop RNAs may be ancestral regulators. I consider the RNA (ribozyme) world scenario from the perspective of addiction modules and cooperating quasispecies (i.e., subfunctional agents that establish group identity). Such an RNA collective resembles a "gang" but requires the simultaneous emergence of endonuclease, ligase, cooperative catalysis, group identity, and history markers (RNA). I call such a collective a *gangen* (pathway to gang) needed for life to emerge.

Keywords: virus; evolution; group selection; cooperatively; origin of life; gangen

The ever-present virosphere

All living habitats (including prebiotic ones) have and must operate in a virosphere (a network of infectious genetic agents). The authentic survivability of life must also be measured in the virosphere. Although the realization of the ubiquity, scale, and diversity of the virosphere is a rather recent development, it still identifies a fundamental feature applicable to all life. However, most experimental paradigms seek to eliminate or have ignored viruses. For example, when we clone *Escherichia coli* free of temperate and lytic phage or when we establish a sterile mouse colony free of all the usual persistent viruses, we create a misleading virus-free habitat for the survival of life. I suggest that, to better understand the origin and evolution of life, we must instead adopt a virus-first perspective. In such a perspective, the persistence of virus information becomes key. It is through such a perspective that we are led to think of consortia, not clones, as the more fundamental features of life. This perspective also includes the characteristics of both competition and symbiosis, as viruses are inherently symbiotic parasites that can also compete fiercely (killing or protecting host).

Dark matter of the virosphere: persistence

Viruses can transfer genes between themselves and their host. However, it is much more common that they transfer virus-derived information to their host, as demonstrated by any metagenomic analysis.[1] Stable transfer of the entire information content of the virus is a definition of virus persistence. But even transfer of partial (defective) viral information can lead to virus–host persistence at a population level. Thus, viruses often stably colonize their host and persist (symbiosis). Viruses can also kill their host, but persistent viruses can

doi: 10.1111/nyas.12565

protect their host from the very same (or similar) virus. It is asserted here that, together (virus killing and virus protection), we see a truly creative and cooperative force in the evolution of life that is fundamentally both symbiotic and competitive but that affects populations, not simply individuals (owing to virus transmission). Since virus persistence is exceedingly common but usually a silent state, it represents a large but mostly unnoticed force in evolution—the dark matter of biology. In addition, viruses will spontaneously parasitize themselves by generating defective virus. Such defective parasites of viruses appear to be "junk" to most observers, and are indistinguishable in function or consequence from what we call *transposons*. The persistence of such defective parasites can, nevertheless, also provide protection against virus killing. Thus, the persistence of virus information can promote virus–host survival yet also provide survival advantage to populations that retain the ability to produce lytic virus and kill competitor populations.[2] This virus–host dynamic is an ancient, ongoing, and inherently symbiotic force in evolution. Together, these opposing functions of protection and killing provide and define an acquired host group identity. Such group identity allows subfunctional consortia to attain a combined greater competence and operates via a dynamic, consortial, and history-dependent mechanism. Thus, a main objective of this communication is to experimentally justify and further clarify this perspective and to extend this thinking into the RNA world. As it fundamentally involves nonlinear consortia (networks), understanding it will be counter-intuitive and difficult. This perspective depends on an ever-present virosphere, which provides a creative–destructive combination force for the origin and development of life. However, the real power of the consortial action of genetic parasites will be best understood through the action of stem-loop RNA. These simple ancestors to life and viruses are particularly competent to function as consortia.

A main strategy for the persistence of parasitic information is the addiction module

Addiction modules were initially identified through the study of episomal DNA phage. Addiction modules define a core strategy for virus persistence, but also lead us to understand how viruses provide a path to cooperation via the combination toxic destructive (lytic) action and counteracting protective (immune) action. P1 is a stable episomal prophage of *E. coli* that is ubiquitous in wild isolates and has long been studied for its ability to interfere with infections by other phage.[3] Initially, P1 was considered as a plasmid, but its recognition as a persistent and lysogenic phage was soon realized. The mechanism for this stability was first discovered by the Yarmolinski group at the National Institutes of Health in the 1990s after many years of study of postsegregation killing.[4] The virus stability is mediated by an addiction module that comprises a stable protein toxin and a less stable protein antitoxin that are coregulated and act in coordination.[5–7] Loss of plasmid (virus) during cell division into daughter cells leads to the killing of the "cured" cells by the stable toxin. The ability of the P1 addiction module to induce postsegregation killing, however, also involves the cells' own programmed cell death systems, such as the mazEF toxin/antitoxin gene pair.[8] Indeed, it has been proposed that this self-killing (programmed cell death), besides insuring maintenance of P1 prophage, can be a defense mechanism that inhibits the lytic spread of P1.[9] Such observations led me to generalize the concept of P1 addiction from a process that insures the specific maintenance of P1 and promotes its survival to one in which combinations of persisting cryptic prophage (often hyperparasites) will together provide resistance of the colonized host to a diverse set of viruses, such as those in the ever-present virosphere.[10–13] The presence of P1 will kill cells infected by other phage. P1 itself can be colonized by IS2, which can interrupt addiction modules and change the host–virus relationship with other viruses.[14] Interestingly, similar insertions of IC family restriction systems into P1 can also be seen as linked to the horizontal spread of DNA restriction systems.[14] Since such states involving genetic parasites being colonized by other genetic parasites are very common and they can significantly affect the relationship of the colonized host with other viruses, I have previously called this a *hyperparasite* colonization that provides a network-based virus–host system affecting its viral ecology.[2,10] This raises the interesting question of how an addiction system (like P1) might be modified by yet further colonization. Clearly, cell death would need to be prevented by new colonizers. I have argued that these viral (and subviral) agents are the

principal mediators of acquired host group identity. But besides affecting host and group survival in the virosphere, persisting viruses can also often be the source of novel host molecular systems.[15]

Collective actions of dispersed defective viruses: protective and destructive

The discovery of addiction modules and their relationship to persisting viruses has mostly been in the context of bacterial double-stranded DNA viruses. From this, we see that the collective action of dispersed seemly defective (cryptic) viruses can provide specific functions (such as mobilization and network control). But, as asserted in my introduction, a host cell population that is persistently colonized by such a controlled or cryptic virus set will also be able to produce or resist the action of the equivalent lytic virus(es). Thus, a competing identical population of host cells that are not persistently colonized will be susceptible to lysis when it becomes exposed to populations of cells that are persistently infected. This is essentially why a lysogenic strain of bacteria will lyse an identical bacterial strain that is not lysogenic when the two populations are mixed. The lysogenic strain can "reach out" and kill its otherwise identical neighbor via transmissible virus. Since this can happen with episomally persisting agents, it need not directly involve the host DNA genome (it can be epigenomic). The history of virus exposure and colonization will therefore determine whether a specific host population will be lysed or resist a particular virus. This history is stochastic, however, and cannot be predicted. However, to continue to favor survival of the virus-persisting population, these cells must maintain both the capacity to resist virus and the capacity (or reside in a habitat with the capacity) for the production of lytic virus. Hence, viral junk must remain in the virosphere. This is, thus, a virus addiction module, and both protective and destructive functions are required to favor the survival of persistently infected populations, especially in a diverse virosphere. In this way, viruses promote the emergence of a group identity in their host. The bacterial identity will be very much determined by its colonizing set of genetic parasites. Although such assertions seem broadly important, in my judgment what is even more broadly significant is that this situation defines a strategy by which a collective of subfunctional and opposing agents can participate in the genesis of a collective function and group identity. This requires a coherent network that is inclusive of opposing functions (various toxin–antitoxin (TA) sets), but favors persistence of the new parasite-derived information. Cryptic prophages are indeed the main source of new TA sets in prokaryotes, but such new sets must counter or interact with existing TA sets to persist. Such a strategy should also apply to various RNA agents thought to have participated in the origin of life. What addiction modules and group identity allow us to explore is how a collective of subfunctional RNA agents might have been able to become a coherent group that has both function and a TA system needed for group identity. RNA is the crucial agent population to understand, for it underlies the origin of life and the regulation of much complexity in higher organisms. Can virus addiction and group identity help explain creative RNA functions?

Group identity as fundamental

The existence of virus-mediated group identity has much deeper implications: it can also allow us to propose a role for viruses in the origin of life itself. How might a virus lifestyle predate the origin of life? Is a virus not a parasite of a ribosome-containing cell, therefore only able to emerge after the emergence of cells (and ribosomes)? But a virus, at its core, is a molecular genetic parasite that can even parasitize other virus systems. This means any prebiotic replicator system can and likely will be susceptible to virus emergence and colonization (before the evolution of ribosomes). However, the virus-mediated addiction module leads us to also think how opposing functions might emerge and support consortia functions or group identity in early life. This can provide us with a major insight! If viruses can function as a consortium, then this might provide mechanisms from which consortial functions themselves could emerge in prebiotic life. Genetic parasites can act as a group. But for the groups to be coherent, they must attain group identity via an addiction strategy.[2] However, in contrast to traditional and linear thinking, all these features (group identity, addiction modules, regulatory complexity, network emergence, host–virus ecology, host–host competition) are fundamentally interlinked and consortial. They are inherently network phenomena. They cannot be teased apart to define a single and linear logic as is currently accepted for individual fittest-type selection. For example, one cannot

understand the origin and function of the viral toxin (including lysis) without also considering the viral antitoxin (the opposition of self by persistence or defectives). They must be considered to have emerged together. Group identity (network membership/security) is asserted to be a fundamental (but generally ignored) feature of all living systems, even prebiotic systems. Group membership cannot be understood as a linear system (or a linear network). Such a requirement will almost certainly confuse us. Our very language compels us to think in linear terms. Thus, we readily accept the linear thinking of individual fittest-type selection, not a gang-like action, as being the most powerful and creative force. But if life is fundamentally consortial, inclusive of opposing actions, such linear thinking will fail.

Eukaryotes host RNA agents

In the genomes of eukaryotes, RNA agents (retroviruses and retroposons) are much more diverse, numerous, and dynamic (relative to prokaryotes), and provide multiple levels of regulatory complexity. We have recently come to realize that transcription of such retroposon sequences (previously considered junk) is abundant and often produces noncoding RNAs with stem-loop regions (see the ENCODE project[16]). It is such RNA that appears closely involved in complex multicellular identity. However, as I will now present, retroviruses are the major initiators of retroposon-mediated changes in eukaryotes, and the fitness of retroviral RNA (like all RNA) is fundamentally consortial or quasispecies (QS) based. However, this is not the QS as most understand it (based on error and master fittest type). It is, instead, a cooperative and counteractive QS that supports group identity.[17] It is now asserted that life will only emerge from consortial systems with group identity. But early RNA-based life forms, like RNA-only viruses, cannot persist as DNA. The original RNA world must have persisted either as a dynamic RNA population or as a sequestered (static) RNA population. One present day example of dynamic RNA persistence is hepatitis C virus (HCV). HCV is thus presented below as an exemplar of QS-mediated group identity.

Cooperative QS, group identity, and the HCV exemplar

The capacity of a particular RNA virus QS to outcompete and displace any former QS of the same virus was initially observed on the 1990s.[18] The QS collective is thus more fit than any individual member, owing to complementation and sharing of gene products.[19] In 2006, it was further observed that, to attain disease-associated fitness in an mouse model, an RNA virus (polio) needed to generate a virus diversity that acted cooperatively[20] (for review see Ref. 21). Such cooperative behaviors led to our proposal for cooperative quasispecies (QS-C).[22] Inherent to this QS cooperation, however, is competition and interference within the population of minorities.[21] To maintain replication, any single RNA must be coherent with both the cooperative and interfering features of the various subpopulations. It must fit in with and be maintained by these combined toxic and antitoxic functions. Like an addiction module, such features should lead a QS to also provide group identity. If this line of reasoning is correct, we should readily see experimental or naturally occurring results that show QS-based competition. Indeed, there is strong evidence for this. The best-studied RNA virus QS (not persisting as DNA) is HCV. [23] Here, there are distinct HCV clades (QS) that can preclude one another *in vivo*. Through both blood donation[24] and HCV-infected liver transplants into a distinct HCV-infected recipient,[25,26] clade displacement has been observed. Only one strain will survive these mixed states.[27] Thus, HCV replicons show a high degree of competition.[28] Such QS-based displacement can to be very rapid and has been observed within 1 day after transplant.[26] Such behavior is consistent with QS-based preclusion (group selection). Some *in vitro* experiments have evaluated the mechanism of superinfection exclusion. observing that it can occur via RNA replication.[29] Interestingly, selection for HCV variants that can overcome such exclusion can result in alterations to the 3′ UTR region, apparently operating via RNA replication.[30]

Eukaryotic networks, regulatory complexity, and small RNA

Retroviruses also clearly generate and evolve via QS.[21] But, in contrast to the RNA-only viruses (i.e., HCV), retroviruses persist as and are copied from DNA. Importantly, retroviruses have also provided a large amount of genomic DNA sequence (especially from their long terminal repeats (LTRs)), as found in most eukaryotes.[31] If such genomic endogenous retroviral (ERV) sequences are also produced by

QS-mediated evolution, then their involvement in the formation of new or edited networks regulating host functions might be understood as also resulting from a consortial QS RNA-based process with inherent group coherence. Indeed, understanding the origin of transposable RNA-based networks (and the needed network security) has always been challenging from a Darwinian perspective, as networks do not fit into tree-based analogies.[32–34] In addition, it appears that various small noncoding RNAs often participate in multitask networks, and such RNA tend to be transcribed from junk retrotransposons.[35–38] ERVs appear to have been active editors of the human genome. There are about 330,000 solo LTRs in human DNA,[39,40] each of which must have initially corresponded to an intact ERV (∼10 kb) subsequently lost by deletion. This means that during our evolution, 3.3 gigabases of human DNA (current size of our genome) was once retrovirus.[41] Such LTRs are highly involved in the emergence of new regulatory networks, such as in the origin of the placenta:[42–46] it is estimated that LTRs contributed to reregulating about 1500 genes needed for the placenta to emerge.[47,48] Even more fundamental, in the African primates, 320,000 LTRs altered p53 binding sites; even such a basic control network can be edited by ERVs.[49] These primate p53 network changes, however, also relate to (cooperate with) other networks, such as brain-specific microRNAs[50] and alterations to DNA methylation involved in controlling SINE-[51] and Alu-derived transcription.[52] Thus, these LTR-mediated changes show complex interconnections to various other networks.

LTRs: basal importance of stem-loop RNA and RNA virus regulation

Core functions of retrovirus LTR regulation are mediated by various stem-loop RNA structures (including tRNA primers) found in both the 5′ and 3′ ends of retroviral RNAs, which provide replication and packaging identity (for an HIV-1 example, see Ref. 53). Thus, LTRs are providing a large set of potentially regulatory (and self-identifying) stem-loop RNA information content to their host genomes. Indeed, stem-loop RNA structures are core identity regulators for most, if not all, RNA viruses, including satellite tobacco mosaic virus RNA, the simplest of all positive-sense stranded RNA viruses.[54] Such noncoding RNA structures also show crucial and cooperative and context-dependent long-distance interactions.[55] Given that such stem-loop structures are even crucial for the function of viroid RNA as a hammer-head ribozyme,[56,57] it has been proposed that stem-loop RNAs are the likely ancestors to all RNA-based life forms, including viruses.[58] However, all prior proposals regarding a possible role of stem-loop RNAs in the origin of life (and viruses) have assumed that Darwinian evolution (individual fittest type) must originate the selective process. In contrast to this, QS-C–mediated evolution would provide a crucial capacity for cooperation and the emergence of group identity early on. As I assert below, a living RNA network will emerge only after a population of subfunctional RNA agents attains both cooperative function (replication) and a collective group identity through the action of linked and coherent positive and negative functions (TA sets). It will now be argued that the ligation and endonuclease activities of stem-loop ribozymes provided the core linked and opposing TA functions needed to initiate and define RNA-based life.

The RNA *gangen* hypothesis

In the QS-C concept, it was argued that agent diversity (not errors) was essential for the capacity of a collective of RNA agents to function cooperatively.[22] Thus, a society of subfunctional RNA agents would be the expected predecessors of RNA-based life. And the type of Darwinian (individual fittest type) selection that is now so familiar would not emerge until DNA emerged to provide genomes. DNA essentially functions as a habitat for the living RNA collective.[59] But before DNA, such an RNA collective must have been able to hold itself together in order to function as a selected population. Although RNA could be stabilized in a viscous complex (similar to a nucleolus), the instability of RNA would most likely require a dynamic state with ongoing replication to provide stability. If so, in order to behave as a population or group, a QS-C must have some process that only promotes members to replicate. Fundamentally, this property is inherent in QS-C behavior. A robust population coherence would require some process that prevents the occurrence of extreme behavior such an overly potent individual defectors or overly active individual selfish replicators. Self-parasitizing defectives can provide this internal control. Thus, essential minorities and defectives provide regulatory functional diversity.

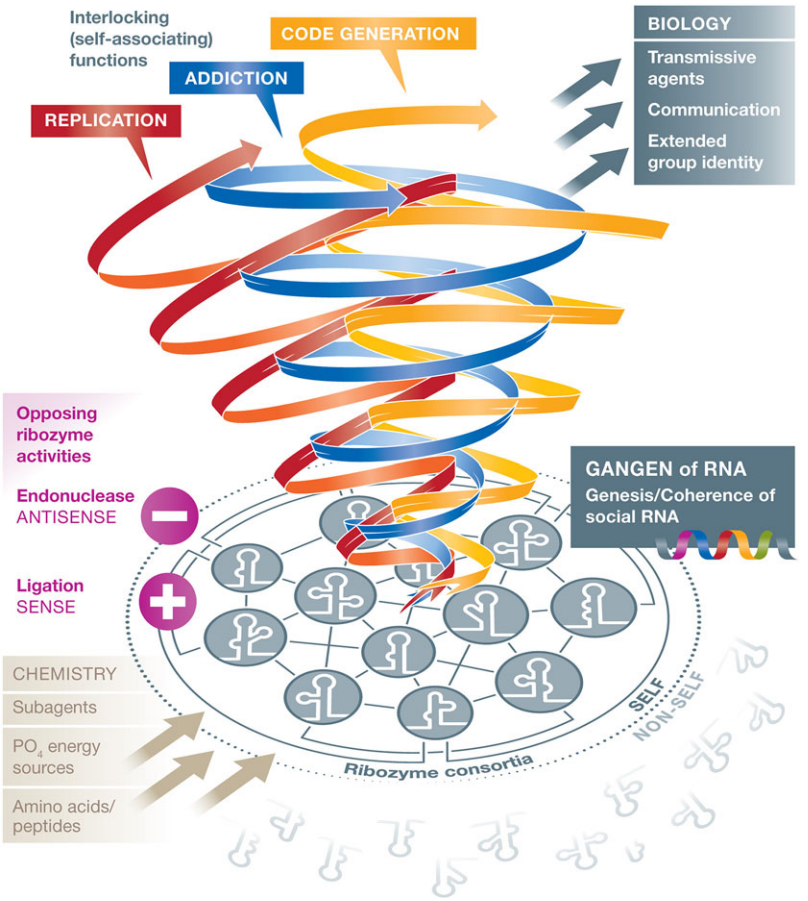

Figure 1. The RNA gangen hypothesis: group identity and cooperativity of an RNA collective that requires opposite functions for the genesis of life (social behavior of agents).

Minorities in the population will also retain memory (learning) from past group-selection events. Minorities must be members of the group identity, and the group must also oppose nonmembers (such as other QS-Cs, as observed with the RNA viruses). Negative (toxic) functions will thus be required for this group opposition, but they are also likely to emerge from cooperative group behavior of subfunctional agents. To attain coherent group behavior, a cooperative system must also attain coherent group identity (via TAs). Any TA function participating in group behavior would need to be coherent with the rest of the TAs already found in the population. Thus, TA coherence (network membership) is required. Regarding the emergence of an RNA ribozyme-based living collective, it will need both opposing ribozyme activities—replication (ligation) and endonuclease—to emerge. Thus, the collective must initially emerge as a complementary collective (not an individual), with group identity mediated by cooperative subfunctional agents that together provide both the replication (positive) and endonuclease (negative) functional features of an addiction module. This means there was no ancestral individual fittest type; the collective was always a dynamic nonlinear network with clearly defined membership (security, immunity) that depends on competition, cooperation, opposing functions (antisense), and the history of RNA agent colonization and their corresponding TA sets, which must attain coherence and provide coherent communication, code, and group identity. In the origin of life, this was mediated mostly by a collective of stem-loop RNAs. I call the hypothesis for the emergence of collective RNA-based life the *gangen* hypothesis, as shown in Figure 1. All the features noted above are included in

the diagram. The word *gangen* was an early Nordic term applied to pathways (gangways) but led to descriptions of collectives (gangs) with clear collective functional abilities and group identities. Here, it also describes the emergence of group membership and the collective living functions of the RNA agents. Membership is not a by-product of individual selection (kin selection), but enforced by the collective. The collective, because it is dynamic and depends on diversity, will also inherently retain memory of its history. Emergence of a gangen is therefore not a simple, chemically predetermined event. It depends on stochastic and historic agents that are able to join the collective and add and edit code and its meaning (use). This provides a distinction between the principles of chemistry and biology (a living collective with history and communication). Clearly, there is more to understanding the emergence of life than this hypothesis can account for. For example, the needed physical containment of the QS population (such as viscosity or membranes); the sources of metabolic energy, substrates, etc.; and the role of amino acids as catalytic RNA primers or markers of replicator identity are not addressed and will not be considered here. But there is an additional feature that should be emphasized, for it relates to the origin of the virosphere. This issue reduces to the idea that a collective of agents (RNA) with inherent toxic and antitoxic features should be able to transmit (communicate) these agents and their features to nearby competing populations (via simple diffusion). Such a transmission is an essentially viral (or viroid) feature. But in so doing, these agents strongly favor the survival of the population with the appropriate addiction modules that will inhibit agent toxicity (prevent lysis through defective code) and allow persistence of the transmitted agents. This is survival of the persistently colonized (infected), which is inherently symbiotic. It also promotes increasing complexity (and identity/immunity) of the host collective through new agent addition accretion. Since such a transmission event can also be defined as communication; the emergence of a virosphere must also have been an early and essential event in the origin of life, one that shaped communication of code and created group identity. This concept differs fundamentally from our current (and highly successful) view based on individual type selection of DNA-based organisms. Below, I assemble some evidence from the study of RNA that supports the existence of collective phenomena in the origin of life.

Origin of ribozymes and cooperating stem-loop RNAs: QS-C perspective

Almost all investigations into the role of RNA in the origin of life assume that some form of "master fittest type" of RNA existed that was able to function as a self-copying ribozyme and inefficiently copy itself with a high error rate, as essentially outlined initially by M. Eigen.[60] However, it has been previously noted that group selection of early replicators, along with compartmentalization, might be required to integrate information in the origin of life.[61] However, the QS-C version of RNA selection has not been previously considered. QS-C posits that a subfunctional collective of RNA agents would be ancestral to effective ribozyme-based replication.[22] Recently, there has been an accumulation of experimental evidence suggesting that RNA ribozymes act and emerge from collectives that can also spontaneously form networks. Very small hairpin ribozymes are known to have catalytic activity.[62,63] Populations of evolving ligase ribozymes have been maintained by *in vitro* serial-diluted passage.[64] Subsequently, the participation of two RNAs that participate in each other's synthesis from four substrates (cooperation) has been observed.[65,66] Others have also used multiple (up to four) stem-loop ribozymes together to select for combined ribozyme activity.[67] Similarly, four subfunctional fragments of group I intron ribozyme can self-assemble into an autocatalytic ribozyme.[68] It has been established that group I ribozymes must undergo cooperative interactions that depend on native helix orientation to attain their functional three-dimensional folds.[69] Cooperative fragments of RNA replicators have also been observed to spontaneously self-assemble and generate a network with cooperative catalytic activity.[70] In such a network, a single RNA molecule can be multifunctional in an RNA pathway.[71] Together, these results provide strong evidence supporting the cooperative potential of ribozymes. However, in none of these reports or discussions has the issue of network membership (or group identity) been considered. According to the gangen hypothesis, network membership, along with various addiction modules, would also be essential for life to emerge from the RNA world. Accordingly, the ligation and endonuclease activity of ribozymes would

need to emerge together to provide a TA set of functions. The creative and consortial action of RNA populations remains a potent and ongoing force in the evolution of the most complex life forms that continue to inhabit DNA.

Conclusion

It is now possible to consider many other issues from the perspective of QS-C and the cooperative interaction of stem-loop RNAs. For example, ribosomes are composed of a complex set of covalently linked stem-loop RNAs that interact in complex ways to provide it with its core function, the catalytic synthesis of peptide bonds.[72] Given that their individual stem loops appear to have various and distinct evolutionary histories, the ribosome seems to represent a consortium of stem loops that was built up over time.[73] Thus, when it became a resident of DNA, the stem-loop RNA consortium created a stable habitat. But the massive creative power of a cooperative RNA consortium (QS-C) remains crucial for life. QS-C was made known to us only recently by virus evolution (e.g., HIV-1). Its role in the origin of life, the emergence of complexity and the creation of group identity should now receive our combined attention.

Acknowledgments

I thank the Center for Virus Research at the University of California, Irvine, for support in the development and presentation of this manuscript. I also thank Uta Mackensen from European Molecular Biology Organization (EMBO) for excellent graphics work and Vera Kolb for her constructive editorial comments.

Conflicts of interest

The author declares no conflicts of interest.

References

1. Koonin, E.V. 2011. *The Logic of Chance: The Nature and Origin of Biological Evolution.* 530 pp. FT Press.
2. Villarreal L.P. 2005. *Viruses and the Evolution of Life.* Washington, DC: ASM Press.
3. Yarmolinsky, M.B. 2004. Bacteriophage P1 in retrospect and in prospect. *J. Bacteriol.* **186:** 7025–7028.
4. Lehnherr, H., E. Maguin, S. Jafri & M.B. Yarmolinsky. 1993. Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and PHD, which prevents host death when prophage is retained. *J. Mol. Biol.* **233:** 414–428.
5. Lehnherr, H. & M.B. Yarmolinsky. 1995. Addiction protein PHD of plasmid prophage P1 is a substrate of the clpxp serine protease of *Escherichia coli. Proc. Natl. Acad. Sci. USA* **92:** 3274–3277.
6. Magnuson, R., H. Lehnherr, G. Mukhopadhyay & M.B. Yarmolinsky. 1996. Autoregulation of the plasmid addiction operon of bacteriophage P1. *J. Biol. Chem.* **271:** 18705–18710.
7. Gazit, E. & R.T. Sauer. 1999. Stability and DNA binding of the PHD protein of the phage P1 plasmid addiction system. *J. Biol. Chem.* **274:** 2652–2657.
8. Hazan, R., B. Sat, M. Reches & H. Engelberg-Kulka. 2001. Postsegregational killing mediated by the P1 phage "addiction module" PHD-doc requires the *Escherichia coli* programmed cell death system mazEF. *J. Bacteriol.* **183:** 2046–2050.
9. Hazan, R. & H. Engelberg-Kulka. 2004. *Escherichia coli* mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. *Mol. Genet. Genomics* **272:** 227–234.
10. Villarreal, L. 2011. "Viruses and host evolution: virus-mediated self-identity." In *Self and Non-Self.* C. Lopez-Larrea, Ed. Dordrecht: Landes Bioscience and Springer Science + Business Media.
11. Villarreal, L.P. 2009. Persistence pays: how viruses promote host group survival. *Curr. Opin. Microbiol.* **12:** 467–472.
12. Villarreal, L.P. 2009. The source of self: genetic parasites and the origin of adaptive immunity. *Ann. N.Y. Acad. Sci.* **1178:** 194–232.
13. Villarreal, L.P. 2012. "The addiction module as a social force." In *Viruses: Essential Agents of Life.* G. Witzany, Ed. Dordrecht: Springer Science+Business Media.
14. Tyndall, C., H. Lehnherr, U. Sandmeier, *et al.* 1997. The type IC hsd loci of the enterobacteria are flanked by DNA with high homology to the phage P1 genome: implications for the evolution and spread of DNA restriction systems. *Mol. Microbiol.* **23:** 729–736.
15. Chikova, A.K. & R.M. Schaaper. 2005. The bacteriophage P1 hot gene product can substitute for the *Escherichia coli* DNA polymerase III θ subunit. *J. Bacteriol.* **187:** 5528–5536.
16. Harrow, J., A. Frankish, J.M. Gonzalez, *et al.* 2012. Gencode: the reference human genome annotation for the encode project. *Genome Res.* **22:** 1760–1774.
17. Villarreal, L.P. 2009. *Origin of Group Identity: Viruses, Addiction, and Cooperation.* New York: Springer.
18. Clarke, D.K., E.A. Duarte, S.F. Elena, *et al.* 1994. The red queen reigns in the kingdom of RNA viruses. *Proc. Natl. Acad. Sci. USA* **91:** 4821–4824.
19. Novella, I.S., D.D. Reissig & C.O. Wilke. 2004. Density-dependent selection in vesicular stomatitis virus. *J. Virol.* **78:** 5799–5804.
20. Vignuzzi, M., J.K. Stone, J.J. Arnold, *et al.* 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439:** 344–348.
21. Domingo, E., J. Sheldon & C. Perales. 2012. Viral quasispecies evolution. *Microbiol. Mol. Biol. Rev.* **76:** 159–216.
22. Villarreal, L.P. & G. Witzany. 2013. Rethinking quasispecies theory: from fittest type to cooperative consportia. *World J. Biol. Chem.* **4:** 71–82.

23. Martell, M., J.I. Esteban, J. Quer, *et al.* 1992. Hepatitis c virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J. Virol.* **66:** 3225–3229.
24. Laskus, T., L.-F. Wang, M. Radkowski, *et al.* 2001. Exposure of hepatitis c virus (HCV) RNA-positive recipients to HCV RNA-positive blood donors results in rapid predominance of a single donor strain and exclusion and/or suppression of the recipient strain. *J. Virol.* **75:** 2059–2066.
25. Kao, J.H., P.J. Chen, M.Y. Lai & D.S. Chen. 1993. Superinfection of heterologous hepatitis c virus in a patient with chronic type c hepatitis. *Gastroenterology* **105:** 583–587.
26. Ramírez, S., S. Pérez-del-Pulgar, J.A. Carrión, *et al.* 2010. Hepatitis c virus superinfection of liver grafts: a detailed analysis of early exclusion of non-dominant virus strains. *J. Gen. Virol.* **91:** 1183–1188.
27. Fan, X., D. Lang, Y. Xu & A. Lyra. 2003. Liver transplantation with hepatitis C virus-infected graft: interaction between donor and recipient viral strains. *Hepatology* **38:** 25–33.
28. Evans, M.J., C.M. Rice & S.P. Goff. 2004. Genetic interactions between hepatitis c virus replicons. *J. Virol.* **78:** 12085–12089.
29. Webster, B., S. Wissing, E. Herker, *et al.* 2013. Rapid intracellular competition between hepatitis c viral genomes as a result of mitosis. *J. Virol.* **87:** 581–596.
30. Webster, B., M. Ott & W.C. Greene. 2013. Evasion of super infection exclusion and elimination of primary viral RNA by an adapted strain of hepatitis c virus. *J. Virol.* **87:** 13354–13369.
31. Shapiro, J.A. 2005. Retrotransposons and regulatory suites. *BioEssays* **27:** 122–125.
32. Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9:** 397–405.
33. Daly, T., X.S. Chen & D. Penny. 2011. "How old are RNA networks?" In *RNA Infrastructure and Networks.* L.J. Collins, Ed.: 255–273. New York: Springer.
34. Bapteste, E., L. van Iersel, A. Janke, *et al.* 2013. Networks: expanding evolutionary thinking. *Trends Genet.* **29:** 439–441.
35. Mattick, J.S. & M.J. Gagen. 2001. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.* **18:** 1611–1630.
36. Mattick, J.S. & I.V. Makunin. 2006. Non-coding RNA. *Hum. Mol. Genet.* **15**(Spec No. 1): R17–R29.
37. Pheasant, M. & J.S. Mattick. 2007. Raising the estimate of functional human sequences. *Genome Res.* **17:** 1245–1253.
38. Mattick, J.S. 2011. The central role of RNA in human development and cognition. *FEBS Lett.* **585:** 1600–1616.
39. Oliver, K.R. & W.K. Greene. 2012. Transposable elements and viruses as factors in adaptation and evolution: an expansion and strengthening of the TE-thrust hypothesis. *Ecol. Evol.* **2:** 2912–2933.
40. Oliver, K.R. & W.K. Greene. 2011. Mobile DNA and the TE-thrust hypothesis: supporting evidence from the primates. *Mob. DNA* **2:** 8–25.
41. Moelling, K. 2013. What contemporary viruses tell us about evolution: a personal view. *Arch. Virol.* **158:** 1833–1848.
42. Nakagawa, S., H. Bai, T. Sakurai, *et al.* 2013. Dynamic evolution of endogenous retrovirus-derived genes expressed in bovine concept uses during the period of placentation. *Genome Biol. Evol.* **5:** 296–306.
43. Chuong, E.B., M.A.K. Rumi, M.J. Soares & J.C. Baker. 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* **45:** 325–329.
44. Bièche, I., A. Laurent, I. Laurendeau, *et al.* 2003. Placenta-specific INSL4 expression is mediated by a human endogenous retrovirus element. *Biol. Reprod.* **68:** 1422–1429.
45. Harris, J.R. 1998. Placental endogenous retrovirus (ERV): structural, functional, and evolutionary significance. *Bioessays* **20:** 307–316.
46. Emera, D. & G.P. Wagner. 2012. Transposable element recruitments in the mammalian placenta: impacts and mechanisms. *Brief Funct. Genomics* **11:** 267–276.
47. Lynch, V.J., R.D. Leclerc, G. May & G.P. Wagner. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* **43:** 1154–1159.
48. Lynch, V.J., M. Nnamani, K.J. Brayer, *et al.* 2012. Lineage-specific transposons drove massive gene expression recruitments during the evolution of pregnancy in mammals. arXiv:1208.4639 [q-bio.PE].
49. Wang, T., J. Zeng, C.B. Lowe, *et al.* 2007. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl. Acad. Sci. USA* **104:** 18613–18618.
50. Le, M.T., C. Teh, N. Shyh-Chang, *et al.* 2009. MicroRNA-125b is a novel negative regulator of p53. *Genes Dev.* **23:** 862–876.
51. Leonova, K.I., L. Brodsky, B. Lipchick, *et al.* 2013. P53 cooperates with DNA methylation and a suicidal interferon response to maintain epigenetic silencing of repeats and noncoding RNAs. *Proc. Natl. Acad. Sci. USA* **110:** E89–E98.
52. Zemojtel, T., S.M. Kielbasa, P.F. Arndt, *et al.* 2009. Methylation and deamination of CpGs generate p53-binding sites on a genomic scale. *Trends Genet.* **25:** 63–66.
53. Berkhout, B. & J.L. Van Wamel. 2000. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA.* **6:** 282–295.
54. Archer, E.J., M.A. Simpson, N.J. Watts, *et al.* 2013. Long-range architecture in a viral RNA genome. *Biochemistry (Mosc.)* **52:** 3182–3190.
55. Miller, W.A. & K.A. White. 2006. Long-distance RNA-RNA interactions in plant virus gene expression and replication. *Annu. Rev. Phytopathol.* **44:** 447–467.
56. Flores, R., S. Delgado, M.E. Gas, *et al.* 2004. Viroids: the minimal non-coding RNAs with autonomous replication. *FEBS Lett.* **567:** 42–48.
57. Carbonell, A., R. Flores & S. Gago. 2012. "Hammerhead ribozymes against virus and viroid RNAs." In *From Nucleic Acids Sequences to Molecular Medicine.* V.A. Erdmann & J. Barciszewski, Eds.: 411–427. Berlin, Heidelberg: Springer.
58. Briones, C., M. Stich & S.C. Manrubia. 2009. The dawn of the RNA world: toward functional complexity through ligation of random RNA oligomers. *RNA* **15:** 743–749.
59. Villarreal, L. & G. Witzany. 2013. The DNA habitat and its RNA inhabitants: at the dawn of RNA sociology. *Genomics Insights* **6:** 1–12.

60. Eigen, M. 2013. *From Strange Simplicity to Complex Familiarity: A Treatise on Matter, Information, Life and Thought.* Oxford: Oxford University Press.
61. Szathmary, E. & L. Demeter. 1987. Group selection of early replicators and the origin of life. *J. Theor. Biol.* **128:** 463–486.
62. Yarus, M. 2011. The meaning of a minuscule ribozyme. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **366:** 2902–2909.
63. Muller, S., B. Appel, T. Krellenberg & S. Petkovic. 2012. The many faces of the hairpin ribozyme: structural and functional variants of a small catalytic RNA. *IUBMB Life* **64:** 36–47.
64. McGinness, K.E., M.C. Wright & G.F. Joyce. 2002. Continuous *in vitro* evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. *Chem. Biol.* **9:** 585–596.
65. Lincoln, T.A. & G.F. Joyce. 2009. Self-sustained replication of an RNA enzyme. *Science* **323:** 1229–1232.
66. Ferretti, A.C. & G.F. Joyce. 2013. Kinetic properties of an RNA enzyme that undergoes self-sustained exponential amplification. *Biochemistry (Mosc.)* **52:** 1227–1235.
67. Gwiazda, S., K. Salomon, B. Appel & S. Muller. 2012. RNA self-ligation: from oligonucleotides to full length ribozymes. *Biochimie* **94:** 1457–1463.
68. Hayden, E.J. & N. Lehman. 2006. Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem. Biol.* **13:** 909–918.
69. Behrouzi, R., J.H. Roh, D. Kilburn, *et al.* 2012. Cooperative tertiary interaction network guides RNA folding. *Cell* **149:** 348–357.
70. Vaidya, N., M.L. Manapat, I.A. Chen, *et al.* 2012. Spontaneous network formation among cooperative RNA replicators. *Nature* **491:** 72–77.
71. Vaidya, N. 2012. Spontaneous cooperative assembly of replicative catalytic RNA systems. PhD dissertation. Portland State University.
72. Bokov, K. & S.V. Steinberg. 2009. A hierarchical model for evolution of 23s ribosomal RNA. *Nature* **457:** 977–980.
73. Harish, A. & G. Caetano-Anollés. 2012. Ribosomal history reveals origins of modern protein synthesis. *PLoS ONE* **7:** e32776.